



PCT/AU00/00817

REC'D 28 JUL 2000

WIPO PCT

4

Patent Office  
Canberra

7400/817

I, KAY WARD, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 3549 for a patent by JAMES QUEST filed on 19 October 1999.



WITNESS my hand this  
Twentieth day of July 2000

*K Ward*

KAY WARD  
TEAM LEADER EXAMINATION  
SUPPORT AND SALES

**PRIORITY  
DOCUMENT**

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

AUSTRALIA

---

Patents Act 1990

---

PROVISIONAL SPECIFICATION

APPLICANT: JAMES QUEST  
NUMBER:  
FILING DATE:

Invention Title: SPEECH RECOGNITION SYSTEM AND METHOD

The invention is described in the following statement:-

---

VRT has many significant applications including speech to text conversion for sound activated word processing, natural speech synthesis for messages (the so-called Talking Timetable), voice activated hands-free control such as for example can be used for vehicular and appliance control, voice activated control of bio-  
5 medical devices for the disabled, name dialling for telephones etc.

However even with modern computing systems; the development of VRT has lagged its potential, particularly in relation to SE.

It is believed by the inventor that VRT associated with SE has not achieved its potential and advanced to the extent that technology would otherwise permit  
10 because modern SE is traditionally regarded as a stress timed language and is analysed for VRT purposes in accordance with the classical Saussurean universal language sign and the Applied Linguistic general definition and other theories of language that are based upon them.

This present invention is based on the inventor's understanding that rather  
15 than being a stress timed language, SE has an analytic (hereafter called analytical) phonology in which words, as defined, have two values or orders of signification.

The first value is a standardised or fixed phonetic value and entity which is set by convention, such for example, as defined by the phonetic entries of  
20 individual words found in dictionaries using the International Phonetic Alphabet (IPA) and other systems of phonetic notation such as those used by American dictionaries. The word's first value possesses an agreed upon meaning.

The second value is a variable phonetic value and identity. The second value is variable in time and sound qualities and is defined by the individual  
25 speaker at the moment of utterance. The word's second value possesses a meaning that emanates from its first value but which is also relative to the circumstances of the word's immediate "real life" context and the word's place within the flow of connected speech.

### 30 **Summary of Invention**

The present invention aims to provide an alternative to known speech recognition systems and methods.

---

It is also preferred that the variable indicators include the facts of the immediate context pertaining to the words in the flow of connected speech.

In a preferred embodiment of the invention the method further includes:-

recording speech spoken by a speaker;

5 indicating to the speaker the meanings of the variable indicators of the recorded speech, and

designating or affirming the meanings of the variable indicators indicated to the speaker.

10 It is preferred that the method also includes storing data representative of analysed words for which the meanings of the variable indicators have been designated or affirmed.

In another aspect this invention resides broadly in a system for recognising speech consisting of words having syllables and phonemes, the system including:-

15 recording means for recording speech spoken by a speaker;

means for assigning a first order of signification to a word, the first order of signification including standardised indicators having agreed meanings independent of the speaker;

20 means for assigning a second order of signification to a word, the second order of signification including variable indicators having meanings which are generated by the speaker and are dependent on the context of the word in the flow of connected speech;

indicating means for indicating to a speaker the meanings of the variable indicators of the recorded speech, and

25 designation means whereby a speaker designates or affirms the meanings of the variable indicators indicated by the indicating means.

It is preferred that the system also includes analysing means for analysing the word in accordance with the first and second orders of signification.

30 It is further preferred that the system includes storage means for storing data representative of analysed words for which the meanings of the variable indicators have been designated or affirmed.

In a further aspect this invention resides broadly in a method of teaching how to speak a language, the method including:-

---

FIG 3 is a flow chart of a method of recognising speech in accordance with the invention, and

FIG 4 is a schematic block diagram illustrating a system for recognising speech in accordance with the invention.

5

### **Description of Preferred Embodiments of Invention**

Before providing a more detailed description of the preferred embodiments of the methods and description of this invention, a description of the inventor's understanding of underlying principles will be given, first at a more general level and then in summary.

10

#### **- Synthetic and Analytical languages**

The key languages that helped to shape modern English are what linguists call synthetic languages. Modern German, for example, is a synthetic tongue, as was Latin.

15

A synthetic language is one that possesses a complicated system of grammar dependent upon the use of inflections. Inflections are word endings or affixes that denote things such as gender (inanimate objects having a sex), case, voice, verb tense and number. In spoken German, the more complex an idea to be expressed, the more inflections are required thereby producing words of many syllables. The modern Germanic tongues thus continue to use inflections as an integral part of their grammatical systems.

20

By contrast, Modern English is called a syncretic or analytic language (hereafter referred to as an analytical language or an analytical system).

25

An analytical language is oppositional to the synthetic languages in its core philosophy and drive. A key feature of an analytical system is that it does not depend upon inflections for expressing complex ideas and meanings. The system of inflections was largely abandoned sometime during the period currently referred to as Middle English. Instead an analytical language follows the reductionist path in seeking to express more complex ideas and meanings with shorter words, through the use of simpler grammatical forms and structures, through the flexibility of the function of words that may assume various parts of speech, a flexibility which also enables the constant creation of catch phrases, context-specific jargon,

30

### - The Analytical Phonology

Because modern SE is not a stress timed language as are those Germanic tongues governed by a stress timed phonology, it is beneficial to consider SE as being represented by a new category of phonological system - the analytical phonology. Another example of an analytical language system is written Chinese which uses ideograms and not words composed of characters from an alphabet.

The new category of the Analytical Phonology redefines the nature, general properties and general principles of the English language phonological system that apply where English is spoken as the sole mother tongue of the majority, ie Australia, Canada (other than Quebec), New Zealand and the United States of America - countries which were colonised and populated by England and in Great Britain and Ireland.

The primary principle and purpose of an analytical language system, in either written and spoken language, are its indomitable urge to express increasingly more complex meanings and ideas while at the same time reducing the actual number, form and length of words as well as simplifying the grammatical structures to communicate these meanings. It is in modern SE that the analytical principles and purpose of the language are at their most potent.

Put simply, the analytical phonology of modern SE constantly seeks ways and means to say more with less.

### - Stress in Spoken English

The traditional classification of English as a stress timed language is believed by the inventor to be flawed because connected English speech is not governed by any of the principles of stress timing. This is because stress in SE is fixed to no one stable timing principle but is deregulated and highly unpredictable in its distribution and its timing formations in connected speech. Stress in SE follows no one phonological principle of timing and is capable of adopting either syllable or stress timed patterns and formations, or any other irregular patterns or formations obtainable in speech.

Consequently, within the definition and parameters of the free syllable, the protected syllable and the restricted syllable, the gradations of stress in speech are otherwise freely transferable between phonemes (particularly the vowel sounds), syllables and words within utterances.

There are many and various gradations of stress obtainable in natural connected speech, not merely weak and strong. In general, stress in SE cannot

The basis of variable stress is rational and logical and is the consequence of the language system being governed by no one fixed principle of timing.

Variable stress can be conceptualised as a phonological freedom rather than an imposition or restriction, and which is granted rather than imposed from within.

This freedom becomes manifest and widely exploited in everyday speech. Variable stress and its consequences on the phonological and suprasegmental features of everyday speech allows individual speakers enormous freedom to engineer the sound qualities and patterns of speech enabling them to signify meanings and ideas in tremendously creative, idiosyncratic and inventive ways.

### - Sound and Timing Variables in Spoken English

The key phonological functions of SE dependent on variable stress (hereafter called the key phonological functions or the key phonological features) include variable timing and duration, variable rhythm and tempo and variable pitch and tone. Other suprasegmental and prosodic features of speech effected by variable stress (hereafter called other suprasegmental features) include speed of delivery, volume, word linking/not linking, enunciation, pausing and phrasing.

All of the above key phonological features and phonological functions and the other suprasegmental features of everyday connected English speech are inextricably linked to variable stress. Stress is well understood to have a commanding role to perform in the organising functions of the phonological system. However because stress in SE is variable, the interconnected phonological functions are also variable.

Because stress is variable the role of the other suprasegmental features of speech becomes important. These features generally are variable and/or optional. Their effects on the phonological functions in connected speech greatly heighten the system's overall capacity and scope to accommodate greater sound and timing variations and contrasts which variable stress enables.

All of the sound functions and features of everyday speech are capable of variation within a wide and legitimate parameter of sound & timing variability that the English phonology permits and encourages and which, crucially, the English language phonetic system tolerates and accommodates.

The factors of sound and timing variability in connected speech work at the basic level of the phoneme, the syllable and the word and at the more general level of spoken phrases and connected speech.

A fundamental axiom of the analytical phonology is that the native speaker is free to apply any of the many variable phonological functions and suprasegmental features in the pronunciation of phonemes, syllables and words at choice which are obtainable in speech for the purposes of creating further meaning provided that such variations do not negate the agreed-upon meanings of the words.

To these ends and within this fundamental rule, further meaning can be generated when the phonemes and syllables that constitute the word - still an arbitrary symbol possessing a static agreed-upon meaning - are varied in sound and timing according to the speaker's pronunciation and in ways the listener hears and registers as meaningful.

Having regard to the speaker's manner and habits of pronunciation, the variation of sound and timing and the application of the other variable and optional suprasegmental features at the level of the phoneme, syllable and word, cannot but effect the structure, qualities and organisation of spoken phrases and larger passages of discourse as connected speech progresses.

Within phonemes, syllables, words and phrases, and within the general flow of connected speech, specific sound and timing variations (hereafter called variations) will combine to create discernible sound and timing contrasts (hereafter called contrasts). Therefore, in terms of the sound imagery of SE, the following equation applies:-

Phonemes, syllables and/or words + X number of variations =

Words and phrases + Y number of appreciable contrasts in speech.

The potential number of variations and specific combinations of variations capable of creating discernible contrasts within connected English speech is untold. This enables the same English word to obtain a multitude of different sound images in speech, either as a stand alone word or as it finds itself placed within the traffic of connected speech. This is simply because with so many variations, choices and options available the potential combinations of variations able to produce appreciable contrasts are innumerable.

Nonetheless, in monitoring and comparing the qualities of specific phonemes, syllables, words and phrases and in monitoring the overall changes and variations within connected speech (as outlined above) the fluent listener can recognise contrasts and from this distinguish, within the context of the conversation, what is being made prominent or salient in terms of meaning by the speaker.



particular context - we are unable to predict every possible context nor classify the mood, manner of speech and temper of mind of every individual speaker.

5 The onus then falls upon the listener to be able to recognise, register, decode and interpret variations and contrasts for second order meanings. This is something at which the fluent native speaker is proficient. In short, sound and timing variations and contrasts in the sound imagery, or soundscape, of everyday connected English language speech work productively to be the signifiers of second order meanings.

10 This system through which the production of second order meaning in SE occurs may be termed the second order of signification.

The traditional and orthodox linguistic system of producing sounds that signify static agreed-upon meaning, common to all languages, and as conceptualised by Saussure, may be termed the first order of signification.

### 15 - **Analytical Phonology and the Phonetic System**

The analytical nature of the phonology of SE cannot but effect its phonetic system because the language depends upon its phonological system and its phonetic system co-operating in order that the second order of signification may function properly. Moreover, the variable phonological functions and suprasegmental features of everyday SE operate at the level of the phoneme.

The phonetic system of SE differs markedly from the phonetic systems of the non-analytical language systems in numerous ways including the following:-

25 The way words are spelt in English bears no logical correlation, nor readily understood systematic connection, to the way the words are pronounced. The phonetic pronunciation of words is set by common usage and not by the word's spelling. This creates a fundamental dislocation or disjuncture between the written and spoken systems of the English language at the level of the spellings of words and their pronunciations with the onus falling heavily on the practices and conventions of common usage to determine the acceptable and comprehensible pronunciation of words with these practices and conventions often being in conflict and flux.

30 The phonetic system of SE has the ability to tolerate marked and sometimes radical sound and timing variations within individual phonemes which having no one fixed principle of timing the system can permit. The phonetic system will indifferently accept legitimate variations without negating the static agreed-upon meaning of the word and without rendering

35

---

In practice this means words, as phonetically defined, have two values:

- (i) a standardised or fixed phonetic value and identity set by convention and as defined by the phonetic entries of individual words found in dictionaries using the IPA, and other systems of phonetic notation, and
- (ii) a variable phonetic value and identity, in terms of the phoneme's time and sound qualities and value, something that is defined by the individual speaker in the moment of utterance and which is relative to its standardised phonetic value as well as to the circumstances of the word's immediate "real life" context and its place within the flow of connected speech.

The standardised phonetic value of a particular phoneme is the necessary reference point by which the phoneme's second variable value may be recognised, evaluated and interpreted for the purposes of producing second order meanings.

#### - Analytical Phonology and Word Stress

The analytical nature of the phonology of SE also effects word stress. This is because the variable phonological functions and suprasegmental features of everyday SE operate at the level of the syllable and the word. As words enjoy two phonetic values, one being standardised and the other variable, likewise words enjoy two values in the way they are stressed.

The first "word stressing" is standardised by convention and defined in the phonetic entries of individual words found in dictionaries which usually mark which syllables within polysyllabic words customarily assume main, secondary and weak stress.

The second "word stressing" is a variable that is defined by the individual speaker in the moment of utterance and which is relative to the word's standardised stressing as well as to the circumstances of the word's immediate "real life" context and its place within the flow of connected speech.

The system of variable word stressing, as referred to above, operates within certain limits and parameters. These limits are defined by the three kinds of syllables a polysyllabic word may obtain within the flow of everyday connected speech.

The three kinds of syllables are:-

- (i) The distinct or protected syllable.

multiplicity of second order meanings that the speaker wishes to communicate simultaneously.

Brief reference will be made to FIGS 1A to 1F which illustrates the potentially variable word stressings of the word "disappointing" with reference to free, restricted and protected syllables in the analytical phonology.

FIG 1A shows the standardised dictionary stress pattern for the polysyllabic word "disappointing" which consists of the four syllables: **dis**, **ap**, **point** and **ing**.

However the word is subject to much variation in the traffic of connected speech when placed in different phrases carrying different second order meanings.

Thus as illustrated in FIG 1B which shows the potential variation of syllables in connected speech,

- the first syllable **dis**, which dictionaries define as having secondary stress, is a free syllable. The last syllable, the inflection **ing** in this case, is also a free syllable;
- the second syllable **ap**, carrying the reduced sound /ə/, is the restricted syllable, and
- the third syllable **point**, carrying main stress, is the protected syllable.

In the phrase "The result was disappointing" illustrated in FIG 1C, the stressing of the word "disappointing" is as the dictionary would prescribe.

In the phrases "A disappointing result" seen in FIG 1D and "That's disappointing" seen in FIG 1E, the word "disappointing" is still prominent and is stressed variably according to the speaker's construction of the phrase and the particular second order meaning he or she wishes to convey.

In the phrase "Very disappointing indeed" seen in FIG 1F, the word "disappointing" does not obtain prominence in the phrase at all. The speaker chooses to place the vocal emphasis on the words **VERY** and **indeed**. In all examples where the standardised stressing of the word "disappointing" has been varied it is done so in order to generate a second order meaning within the utterance so as to signify, for instance, an emotional or subjective content or meaning.

A word will deviate from its standardised word stressing and assume a variable word stressing and variable gradations of stress and prominence in connected speech for the purposes of signifying second order meaning.

As is the case with phonemic variation, the standardised stressing of a particular word, as the dictionary would define it, is the necessary reference point

words commonly signifies a sense of doubt, incompleteness, or a need to know more on the part of the speaker.

FIG 2B shows the use of a low rising pitch and tone commonly employed by speakers when reading items from a list to signify that the list is not yet complete, and commonly to signify more neutral feedback or mild interest in what is being said.

FIG 2C shows the use of flat or level pitch and tone in speech, commonly employed to signify disinterest, boredom or sarcasm. In general, level tones will also commonly attend routine or impersonal conversational exchanges.

FIGS 2D and 2E show the use of rising-falling and falling-rising pitch and tones, respectively. Both tone patterns are commonly understood to signify greater emotional content and expression attending the speaker's speech, or to signify contrasting or competing meanings, or to signify a change in register, mood or conversation topic on the part of the speaker.

FIG 2F shows the use of the falling pitch and tone in speech, commonly understood to signify completion, such as when reading the final item in a long list. A falling pitch and tone commonly attends information or "wh-" questions (such as What, When, Where etc) which are asked in the expectation that the answer will be readily provided. Falling pitch and tone will also commonly attend declaratives, statements of fact, and mild apologies the speaker is making. In general, a falling pitch and tone will commonly signify completion, and an absence of doubt in regard to the speaker's utterance.

FIG 2G shows the use of a sharp falling pitch and tone which commonly attends stronger apologies, imperatives, firm statements and declaratives. In general, a sharp falling pitch and tone commonly signify finality, certainty, completion and commonly attend utterances that signify there is no doubt at all in the speaker's mind about what is being said.

The use of variable pitch and tone in connected speech by the fluent speaker enables compression of meaning to occur in utterances when the words of the first order of signification communicate one meaning while, at the same time, the appreciable variations and contrasts of pitch and tone within the second order of signification signify other meanings.

Frequently, speakers in posing yes/no questions, which may commonly adopt a rising tone which signifies doubt, may frame their yes/no question in a falling tone, signifying the opposite: certainty.

Reference will be made to FIG 2F which illustrates this point. The question **Did you murder your wife?** combines both a yes/no question, signified by the

could give the phrase a sarcastic meaning. Here, second order meaning, expressed by the voice, is signifying: **You (don't) look fabulous** or **You look horrible**.

The fluent speaker of SE is well acquainted with the everyday practices of using pitch and tone in this way. Within the wide range of variable sound and timing features and combinations of SE available for the purposes of producing second order meanings, speakers are frequently presented with an utterance that economically compresses multiple ideas which express meanings that realise no literal form. Phrases signifying moods, meanings and ideas of, for example, disappointment, anger, annoyance, frustration, pleasure, desire, sarcasm, contempt and so on, may easily be expressed phonologically in tandem with the static agreed meanings of the utterances' first order of signification: that is, the phrase's actual word content and grammatical form. Variable pitch, tone, duration, timing and stress may also be used as signifiers of meaning by speakers within certain regional sub-varieties of SE and within certain idiolects of SE as a kind of 'in-group' speech code for the purposes of identifying one member of a certain speech community or 'in-group' with his or her peers. This is particularly prevalent among younger speakers of SE, such as second generation migrant youth in Sydney, the so-called 'Valley Girls' of Los Angeles and young followers of the Australian TV soap opera 'Neighbours' in the United Kingdom, who have adopted the Australian tendency of the 'mid rising pitch' in natural speech, hitherto unknown in that part of the world.

Thus, subtextual content, emotion, irony, idiomatic codes, complementary moods and meanings can be compressed within the one spoken English phrase without need of the speaker formulating a new spoken phrase in order to express these further meanings.

In this way the analytical purpose of the spoken English language is greatly served.

### - The Analytical Phonology and Phrases and Sentences

Within the parameters defined by the free, protected and restricted syllables, stress is freely transferable between phonemes and syllables within connected speech.

This enables any word within an utterance to obtain prominence, or to stand out in some way within the phrase in the flow of speech. In a particular word in a phrase being made prominent or noticeable in sound and meaning, while the other words in the phrase do not gain as much prominence or do not

obtained and will most commonly introduce highly subjective ideas and meanings into the full quotient of meanings that the phrase can simultaneously support. Additional meanings of this nature would also be highly context dependent.

For example, in particular contexts and depending on the sound and timing qualities of the speaker's sound imagery, especially the contours of pitch and tone and other extemporaneous suprasegmental features that attended the prominent or noticeable word:

"YOU are going!" in placing the focus on the subject could also express the speaker's subjective opinion of the subject (e.g. enthusiasm or disgust);

"You ARE going!" in placing the attention on the subject's intention "to go" may also signify a curt imperative and not simply be a casual affirmation, and

"You are GOing!" in placing the attention on the subject's act of "going" may also express the speaker's personal feelings in regards to the subject's act of 'going' (e.g. regret or relief).

The flexibility of the phonetic and phonological systems of SE enables the variable sound imagery of individual phonemes, syllables and words to gain prominence in speech that signifies meaning. The permissive nature of the sound system of SE empowers the individual speaker with the vocal means by which he or she can choose which particular fragments, parts, segments and passages of speech gain prominence within and between utterances and in what fashion prominence is to be achieved. This phonological freedom allows the individual speaker to productively signify second order meaning to the extent that the speaker's variable sound imagery obtained in the one spoken phrase may support several second order meanings all at once. Although in these cases the second order meanings emanate from the phrase's static first order meaning these can be extra meanings that are semantically in agreement with, independent of, or in opposition to the phrase's agreed upon 'first order' meaning. It need also be noted that such second order meanings are meanings that realise no literal form as they are communicated by variable sound imagery and not words.

This aspect of SE greatly furthers the cause of the analytical language in its desire to express more complex meanings with reduced words, forms and structures.

- The Analytical Phonology: The general principle of variability

variable sound imagery signifying second order meanings could actually be obtained in acts of speech is almost without limit.

It can be said that in principle:

- the standard indicators of speech as understood by the particular community of SE speakers, plus
- the variable sound imagery of these standard indicators of speech generated by the individual speaker and by other individual speakers within and/or without the immediate context, plus
- the variable and unpredictable compass of the immediate context of natural speech at its moment of utterance,

together, furnish SE discourse with the necessary linguistic constituents that facilitate the phonology's systems of internal tension, play and contrast by which variable sound imagery may obtain meaning in speech and upon which the entire system of the second order signification depends.

#### - The Analytical Phonology and the Faculties of the Speaker and Listener in SE

The listener can understand such meanings communicated by sound and timing variations because of the appreciable and audible contrasts they create within the stream of connected speech in ways that the fluent listener notices, registers, decodes and attempts to interpret for meaning. This is a subliminal process as the intention to signify, and the ability to interpret, meanings in the second order of signification need not be conscious.

For the decoding and interpreting of second order meaning the speaker and listener need standardised reference points by which to judge how much and in what ways sound imagery varies. For this they must rely on the standardised pronunciation of words, as the dictionaries or common usage define it, to gauge variations. Listeners and observers will also rely on commonly understood uses of pitch and tone in SE, in the interpretation of more complex second order meanings that the speaker's use of variable pitch, tone and stress may signify. The listener and observer also rely on their own knowledge of the speaker and the context of the conversation in which the variations occur to help him or her evaluate any second order meanings.

Listeners and observers also rely on their own knowledge of English words and grammar and their own proficiency in the English language to help them interpret second order meanings. As second order meanings are often highly subjective so too are their possible interpretation by the listener.

- **Spoken English: a new definition & conceptual framework**

It is believed by the inventor that both Saussure's original construct of the language sign and modern Applied Linguistics long held definition of language are incomplete definitions of the spoken English language and its analytical phonology.

Saussure's concept of the language sign and the orthodox view of human language describe the first order of signification only.

It can be said of a language system with only one order of signification that:  
fixed sounds and phonemes = words which are arbitrary symbols  
possessing no more than their agreed-upon meanings.

Therefore words, as so defined, are the singular currency of linguistic signification available to the speaker for the production of meaning. In some first order only languages further meaning may be obtained through the use of the suprasegmental and prosodic features of speech where these are possible, available and permissible but strictly under the proviso that their use does not interfere with the primary purpose of the one fixed and central timing principle that controls connected speech: to organise and control the flow of connected speech in a way that ensures the agreed-upon meanings of the words are protected and remain fixed and unvaried.

Spoken English, as opposed to the conventional language systems defined by Saussure, possesses active and highly productive first and second orders of signification. Because the language has two orders, and not one, the relationship between the first and second orders of signification alters the definition of the first order.

In the analytical phonology it can be said that within its first order of signification:

standardised sounds and phonemes = words which are arbitrary symbols that possess agreed-upon meanings.

Within the language's second order of signification it can be said that:

in the speakers' use of sound and timing variations and contrasts, variable sound images of standard words are obtained in speech in a way that does not negate the first order signification but, moreover, systematically generates the signifiers of further or new meaning, ie the second order of meaning.

Within the second order of signification the basic signifiers in the production of meaning in speech are phonemes which construct spoken phrases which are the basic units of second order meaning. Hence the SE phrase in everyday



imagery, with these arbitrary symbols awaiting a speaker and a 'live' context for their formulation or reformulation of meaning within the second order of signification.

Among the hallmarks of the modern spoken English language which characterise its underlying logic and unity of purpose and that would distinguish it from the syllable and stress timed language systems are the following:

- the general principle of variability governing the phonology and sound imagery of spoken discourse which invests all words in connected speech with two potential values in meaning: one standardised, one variable;
- diversity: exemplified by the language's copious 'multicultural' corpus, and the many varieties and idiolects of spoken English that the language engenders and invents;
- an expansionist drive: seen in the language's capacity to acquire new loan words from other languages, a ravenous, on-going process that its phonological system readily accepts when exotic words come to be placed into connected speech. Evinced, also, by the ever expanding parameters of legitimate sound variability tolerated in modern and popular varieties over which no official high arbiter of 'correct speech' or language planning presides but which mass media dominates;
- its inexhaustible vocabulary of potential meanings;
- the ability to self-reform and adapt, in that the progressive analytical objective of the language will generally prevail over tradition and convention - in nearly all varieties of SE bar genres such as Received Pronunciation - when the two are in conflict. RP and other recognised stable genres of SE usually change only gradually - or, in the case of genres such as 'Network-American-English' these can sometimes change more rapidly via mass communication - all providing the necessary standardised indicators and conventional modes of speech which furnish discourse with the stable linguistic constituents that make the internal tension, play and contrast, which engenders variable sound imagery, intelligible;
- an ethos of individualism as the system of second order signification is highly speaker-centred;
- an ethos of indifferent equality between its speakers, exemplified by the practical necessity for the uniquely close and co-operative

the syllables' variable indicators which include the key phonological functions of speech (18) such as pitch, tone, duration, rhythm and tempo as well as including other suprasegmental features of speech (19) such as volume, speed of delivery, enunciation, pausing, phrasing and word linking.

5       The words are then analysed (20) in accordance with the first and second orders of signification. Integral to the process of analysis is the defining of the pertinent facts of the immediate context of the words in speech. Factors such as: **When? Where? Why?** is the conversation occurring? **How** is the conversation occurring (mode of exchange: face-to-face, by phone, via  
10   technology etc)? **What** is the social purpose or the business of the exchange and what events have led up to the conversation that have relevance to the immediate context? **Who** are the participants in the exchange, what is their relationship, and what are their manner, mood and temper of mind? This information may be either operator-dependent, or may be generated or  
15   anticipated by the technology within pre-defined parameters and contexts.

In use the method is implemented by recording speech spoken by a speaker (11), analysing the recorded speech as above, and then indicating to the speaker (21) the meanings of the variable indicators of the recorded speech. The speaker then designates or affirms (22) the meanings of the variable indicators  
20   which have been indicated to him or her. Data representative of analysed words for which the meanings of the variable indicators have been designated or affirmed is then stored (23) in storage means for subsequent transformation to another format (24), such as another language, WP text etc.

FIG 4 is a schematic block diagram illustrating a system for recognising  
25   speech such as SE consisting of words having syllables and phonemes. The system has a recorder 31 for recording speech spoken by a speaker 35. Assigning means 32 assigns a first order of signification to a word, the first order of signification includes standardised indicators having agreed meanings independent of the speaker. Assigning means 33 then assigns the facts of the  
30   immediate context, as described above, and a second order meaning to the word/s in speech. The constituent elements of the second order of signification are the same as those described above.

Indicating means 34 indicates to speaker 35 the meanings of the variable indicators in the speech which has been recorded. Designation means 36 are  
35   operable by speaker 35 to designate or affirm the meanings of the variable indicators which have been indicated by indicating means 34 for subsequent

Once the standardised values of English words have been defined by the system with the speaker's supervision and/or input, the system then defines the second value of words, ie the relative or subjective value of the words.

5       The system samples the speech of the individual speaker and operator at its most natural and spontaneous, and in long durations. This is preferably done in software generated contexts such as games, conversations and particular scenarios with which the individual speaker and operator is familiar and orientated. This allows the individual to  
10       produce second order meanings within a defined or known context in an unrehearsed and spontaneous sampling of his or her everyday connected speech, allowing the individual to demonstrate a full and natural range of his or her vocal and pronunciation styles, emotions and registers. The individual's particular and authentic qualities and manner of speech may  
15       thus be recorded.

As the system samples and records the speaker's manner and qualities of speech, it first measures and qualifies his or her specific parameters of sound and timing variability, paying particular attention to the specific factors outlined in the preferred embodiment above.

20       Upon establishing the general parameters of variability evinced in the operator's natural and spontaneous speech, the system notes, measures and defines any variations and contrasts that arise, using the reference point of the standardised sound values.

25       The system notes the individual's particular tendencies, habits and patternings of pronunciation and voice in natural spontaneous speech and then alerts the operator to sound variations and contrasts evident in the operator's samplings of natural connected speech (as compared to the standardised values of words already established).

30       Within the defined facts of the immediate context the system then allows the speaker to define the particular second order meanings that such variations and contrasts may signify. For instance, do certain combinations of sound and timing variations (say the elongating or reducing of the vowel sound in the same word when repeated) mean something? Or do they signify second order meanings which the speaker  
35       had not consciously intended to generate but nonetheless now need to be consciously considered by him or her in light of the facts of the immediate context? Do clusters of contrasts noted by the sampling process in the

---

layers of second order meanings that the prosodic features of the spoken word naturally support and signify in speech.

Allowing individual English language speakers and operators to more effectively communicate with other English language speakers and operators.

Allowing individual English language speakers to communicate with speakers of other languages in both speech and by the written word, such that the second order signification present only in speech will not be lost or nullified when translated or put into the written word.

Allowing individual English language speakers who are hearing impaired or deaf to appreciate second order signification and second order meanings that are encoded within natural speech through the written word.

Allowing individual English language speakers who are sight impaired or blind to appreciate second order signification and second order meanings encoded within the written word and which can be translated back into natural speech.

The present invention can also be used to teach how to speak a language. In use this method includes:-

assigning first and second orders of signification to words in speech, the words having syllables and phonemes, the first order of signification including standardised indicators having agreed meanings independent of the speaker and the second order of signification including variable indicators having meanings which are generated by the speaker and are dependent on the context of the word in the flow of connected speech, and

practicing speaking using different variable indicators in the second order of signification.

When teaching how to speak the English language in schools in English speaking countries information concerning the analytical nature of the language should be included in mainstream English curriculums.

A number of aspects associated with the present invention are emphasised in the curriculums. These include:-

Providing and teaching a proper and appropriate explanation of the nature, practices and principles of the analytical phonology of spoken English.


Explaining to children, from the moment they learn to read and write, the nature of the separation of the language's written and spoken systems at the level of phonemes and their written alphabetical symbols. (Then

standardised pronunciation. Consequently for the computer in existing VRT packages to familiarise itself with the particular operator and recognise the operator's manner of speech, the onus falls heavily on the operator not to vary his or her pronunciation of the sampled words in any substantial way when the system is later in use, lest the machine be unable to recognise the words correctly.

The failure of known VRT systems to come to grips with the second order of signification is believed by the inventor to have prevented computer science from developing the technologies to enable communication with native English speakers at a level above the somewhat robotic-like current speech level.

The invention is also applicable in telecommunications where existing recorded computer generated voices repeat back numbers to the telephone customer. These can be modified to sound more like natural connected speech.


It will of course be realised that whilst the above has been given by way of an illustrative example of this invention, all such and other modifications and variations hereto, as would be apparent to persons skilled in the art, are deemed to fall within the broad scope and ambit of this invention as is herein set forth.



disapPOINTing  
FIG 1A  
 standardised  
 word stressing

*dis*appointing  
*dis, ing* = free syllables  
*ap* = restricted syllable  
 point = protected syllable

FIG 1B  
 potentially variable  
 word stressing



The reSULT was disapPOINTing

FIG 1C



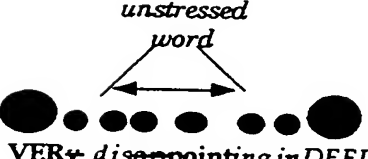
A DISappointing reSULT

FIG 1D



THAT'S DISappointING!

FIG 1E



unstressed  
 word  
 VERY disappointing inDEED

FIG 1F

Falling-rising



Conjunctions: *However, meanwhile,*

*nonetheless* to show contrasting ideas

Some modal verbs: *Could be, should be,*

*can't be, might be, wouldn't be* etc

FIG 2E

Falling



Last item on a list to show completion of ideas: *Sydney, Canberra, Melbourne*

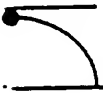
A *wh* question: *Where do you live?*

FIG 2F

Declaratives: *I live in Sydney.*

Apology: *Sorry, Pardon*

Sharp falling



Strong Apology: *Sorry, Pardon*

Imperative: *Now!*

Declarative: *Goodbye.*

FIG 2G

*Did you murder your wife?*

FIG 2H

Speaker A: *I'm going to stop drinking.*

FIG 2I

Speaker B: *I'm sure you will.*

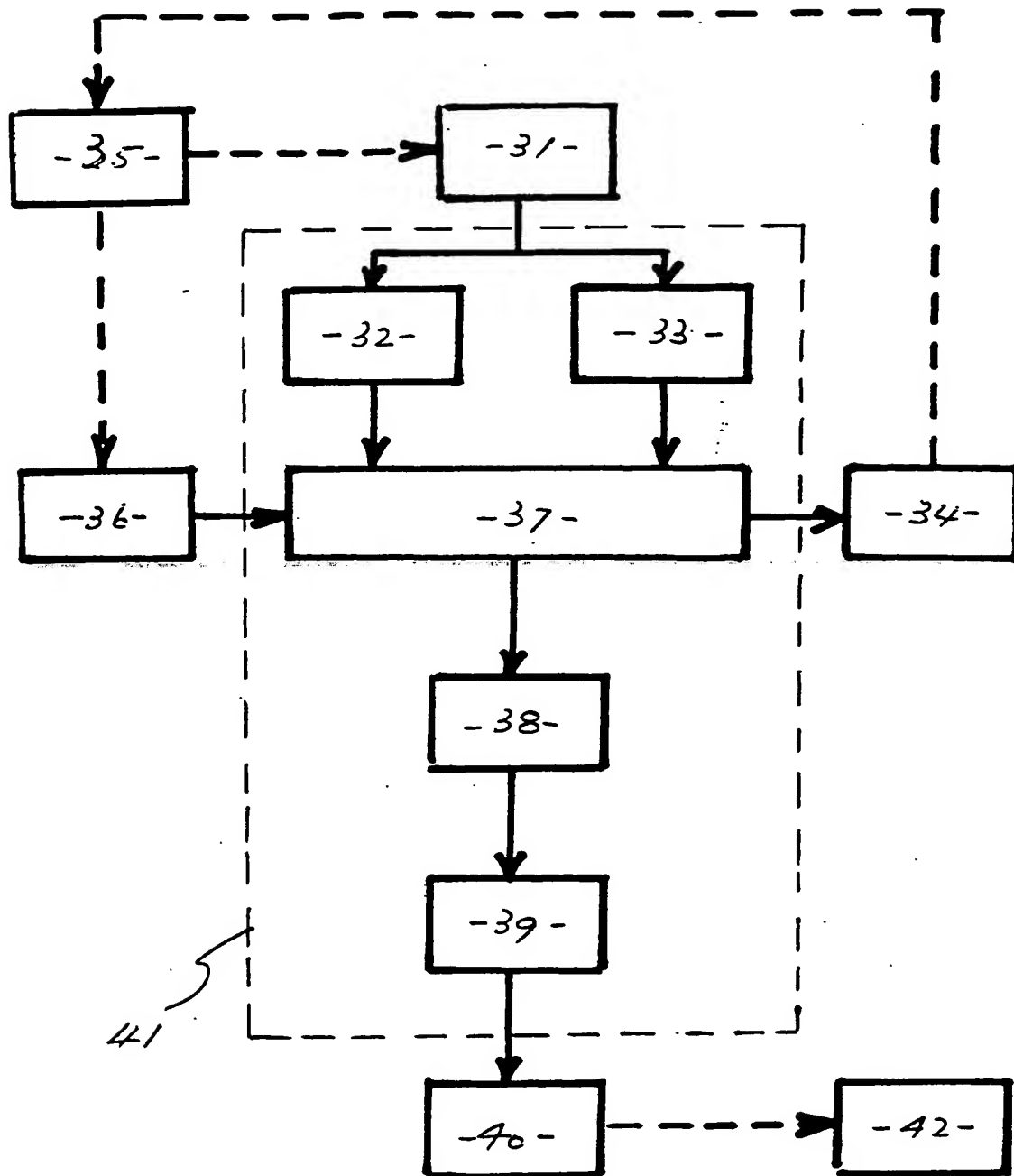
FIG 2J

# KEY

*free syllables*

*restricted syllables*

*protected syllables*

FIG 4